DOCUMENT RESUME

ED 027 022 64 LI 000 045

By-Vinsonhaler, John F.

Improving the Accessibility of Educational Materials: 1. Retrieval of Educational and Psychological Tests. Final Report.

Spons Agency-Michigan State Univ., East Lansing. Coll. of Education.; Office of Education (DHEW), Washington, D.C. Bureau of Research.

Bureau No-BR-5-1144

Pub Date Oct 67

Contract-OEC-6-16-04

Note-42p.

EDRS Price MF-$0.25 HC-$2.20

Descriptors-Automation, Behavioral Science Research, *Computer Programs, Educational Research, Indexing, *Information Retrieval, *Information Services, *Information Systems, Instructional Materials, Tests

Identifiers-*Basic Indexing and Retrieval System, BIRS, Fortran IV

This report summarizes the development, dissemination, and current applications of the Basic Indexing and Retrieval System (BIRS). BIRS is a collection of Fortran IV computer programs designed to allow educational and psychological researchers to use their own locally based computer to construct and maintain the type of information retrieval system which best meets their individual needs. Given a collection of informational elements punched on cards, such as abstracts of research documents, and descriptions of instructional materials, the system produces a variety of information services. These services include the preparation of computer generated books with author/subject indexes and the development of automated reference searches for information elements relevant to specific queries. This report (1) provides sufficient information about BIRS to permit the reader to decide whether the system can be applied to his particular information retrieval problem and (2) describes how the reader may obtain copies of the system and its technical documentation. (Author/JB)

BR 5-1144
LI 000 045
PA 56
PA 64
OE-BR

FINAL REPORT
Project No. 5-1144
Contract No. OE-6-16-04

IMPROVING THE ACCESSIBILITY OF EDUCATIONAL MATERIALS: 1. RETRIEVAL OF EDUCATIONAL AND PSYCHOLOGICAL TESTS

October 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

LI 000 045

IMPROVING THE ACCESSIBILITY OF EDUCATIONAL MATERIALS: 1. RETRIEVAL OF EDUCATIONAL AND PSYCHOLOGICAL TESTS

Project No. 5-1144
Contract No. OE-6-16-04

John F. Vinsonhaler

October 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Michigan State University

East Lansing, Michigan

ABSTRACT

This document summarizes the development, dissemination, and current applications of the Basic Indexing and Retrieval System (BIRS). BIRS is a collection of Fortran IV computer programs designed to allow educational and psychological researchers to use their own locally based computer to construct and maintain the type of information retrieval system which best meets their individual needs. Given a collection of informational elements, punched on cards (e.g. abstracts of research documents, descriptions of instructional materials, etc.), the system operates upon this collection to produce a variety of information services--including the preparation of computer generated books with author/subject indexes, and the development of automated reference searches for informational elements relevant to specific queries. The present monograph (1) provides sufficient information about BIRS to permit the reader to decide whether the system can be applied to his particular information retrieval problem; and (2) describes how the reader may obtain copies of the system and its technical documentation.

TABLE OF CONTENTS

I. INTRODUCTION 1

II. COMPUTER PROGRAMS FOR INFORMATION RETRIEVAL 7

III. THE BASIC INDEXING AND RETRIEVAL SYSTEM (BIRS) 12

A. PURPOSE 12

B. THEORY 14

C. BIRS/I 19

D. BIRS/II 26

IV. APPLICATIONS 30

V. CONCLUSIONS 35

VI. REFERENCES 37

I. INTRODUCTION

The present document is the final report for a research and development project* jointly supported by the Office of Education, Department of Health, Education, and Welfare---and the College of Education, Michigan State University.

The major stated objective of the project was to develop and test a system of Fortran programs for information retrieval which could be widely used by educators (for instructional materials) and by behavioral scientists (for research information).

This principal objective was accomplished by the release of the initial version of the Basic Indexing and Retrieval System (BIRS/I) in December, 1966. BIRS/I satisfied all of the objectives previously stated in the grant proposal submitted to the Office of Education.

BIRS/I is capable of aiding with the basic operations of storage, indexing, and retrieval either with coordinate or weighted indexing. As proposed, BIRS/I was completely tested on a collection of psychological test materials, prior to release.

Once the primary objective of the project was satisfied by BIRS/I, we turned our attention to the attainment of objectives only hinted at by the original research proposal: (1) the development of BIRS/II, a more advanced system capable of handling a thesaurus of synonyms, permitting

---

*USOE Contract #OE-6-16-04, Project #5-1144. Improving the Accessibility of Educational Materials: 1. Retrieval of Educational and Psychological Tests. John F. Vinsonhaler, Principal Investigator.

greater freedom in stating questions, and allowing highly sophisticated methods of automated indexing; (2) the dissemination of BIRS/I and BIRS/II to computer installations interested in using the system; and (3) the application of BIRS/I and BIRS/II to significant educational information retrieval problems.

Surprisingly all three of these hoped-for objectives have been--to some extent--attained since the accomplishment of the contractual obligation by the completion and reslease of BIRS/I in December, 1966. (These additional achievements were certainly not obtained within the limited budget* of the present project. Substantial additional funds were contributed by the College of Education and by other federally funded projects, which required the use of BIRS.) Briefly, the following major developments may be mentioned.

First, the advanced version of the Basic Indexing and Retrieval System (BIRS/II) has been completed and tested on the CDC 3600. BIRS/II provides all of the improvements indicated previously and is now being modified for the U.S.A.S.I. Standard Fortran, to permit easier implementation at diverse computer installations. Complete descriptions for BIRS/I and BIRS/II are given in subsequent sections of the present report. The system and full documentation will be released for national distribution in December, 1967.

*Federal contribution to the present project was about $15,000. This amount was doubled by other contributions.

Second, a dissemination service for BIRS has been established through Educational Publication Services, College of Education, Michigan State University, East Lansing, Michigan. The following table summarizes the major sources of information about BIRS/I and BIRS/II.

Presently, over 100 letters of inquiry regarding BIRS have been answered, about 10 orders for system documentation packages have been filled, and at least five major computer installations are committed to the modification of BIRS for local use.

Third, the Basic Indexing and Retrieval System (BIRS/I and BIRS/II) have been widely applied to diverse information problems ranging from the automation of research libraries to the content analysis of protocols obtained in psychological research studies. Particular applications of BIRS are summarized in final sections of the present report.

TABLE 1: Sources of Information About BIRS/I and BIRS/II

I. REPORTS

Koppitz, E. G. (Editor) "Applications of Computers." Data Processing for Education. Vol. 6, No. 5. June, 1967.

Summary of the BIRS/I System as presented in a report given at the AERA convention.

Millman, J. (Editor) "Personalized Information Retrieval Systems." AERA Educational Researcher, 1967. Supplement Issue: Information Storage, Retrieval and Dissemination.

Brief summary of the BIRS/I system with indication of usefulness of system for development of local, personalized IR systems.

Vinsonhaler, J.F. "BIRS: A System of General Purpose Computer Programs for Information Retrieval in Educational Research" Paper presented at the AERA Annual Convention, 1966.

Analytic review of BIRS/I emphasizing applications to educational research and practice.

Vinsonhaler, J.F. "BIRS: A System of General Purpose Computer Programs for Information Retrieval in the Behavioral Sciences" American Behavioral Scientist, 1967, 10, 12-23.

Analytic review of BIRS/I with emphasis upon social science applications of the system.

Vinsonhaler, J.F. "BIRS: A System of General Purpose Computer Programs for Information Retrieval. Learning Systems Institute Paper No. 39 (Revised). College of Education, Michigan State University.

Definitive research report on the BIRS/I programs including comments on applicability in education and social science, analysis of major weaknesses, and announcement of plans for second version, BIRS/II, to correct faults.

Vinsonhaler, J.F. "Generalized Systems of Computer Programs for Information Retrieval" Learning Systems Institute Paper #55. College of Education, Michigan State University, 1967. (Presented at the 75th Annual APA Convention, 1967).

Review of general purpose programming for IR; Summary of BIRS/II and applications; Relation of the system to other general purpose IR programming developments.

TABLE I (continued)

II. TECHNICAL MANUALS

Vinsonhaler, J.F. (Editor) Technical Manual for the Basic Indexing and Retrieval System. Educational Publication Services, College of Education, Michigan State University, East Lansing, Michigan. Jan: 1967 (First Edition BIRS/I), and January, 1968 (Second Edition BIRS/II, in press).

The technical manual for BIRS/I is now being distributed at cost. It includes chapters providing general reviews of IR, and the analysis of information systems, as well as a technical chapter for operating each component program.

The manual for BIRS/II will be ready for distribution in December, 1967. This edition will include more general reading in IR, special chapters for users and programmers as well as expanded chapters on component programs.

III. SYSTEM PROGRAMS

BIRS DOCUMENTATION PACKAGE

Educational Publication Services, College of Education, Michigan State University, East Lansing, Michigan.

Packages of the materials necessary to implement BIRS at computer installations are being distributed from Educational Publication Services at cost. Each package includes:

(1) Fortran Programs and standard test data punched on cards.

(2) Printed output from compile and execution of all programs using test data.

(3) Three copies of BIRS Technical Manual.

Currently the package includes BIRS/I documentation. After December, 1967, the package will include only BIRS/II documentation.

As noted in the above table, complete technical documentation for the present project is available from Educational Publication Services at Michigan State University.* Therefore, the remaining portions of the current monograph will be devoted to a non-technical report for BIRS. The organization is as follows:

Section II, indicates the research context for BIRS by summarizing the salient attributes of a representative set of similar systems in operation or currently under development.

Section III, provides a summary of the major characteristics of BIRS/I and BIRS/II, including indications of the unique design underlying the system.

Section IV, reviews several representative applications of the system to information retrieval problems in education.

Finally, Section V provides some conclusions drawn from our experience in developing and applying the system.

The goal of the present monograph is to provide a basic source of information about BIRS. Thus, this document should (1) permit an individual scientist or educator to decide whether or not the system may be used to solve his particular information retrieval problem and (2) indicate how complete documentation for the system can be obtained.

*The complete BIRS Documentation Package will be made available without cost to the U.S.O.E. for possible dissemination through the Educational Research Information Center (ERIC).

II. COMPUTER PROGRAMS FOR INFORMATION RETRIEVAL

Information systems are usually analyzed into three basic elements: the source (origin of the information); the receiver (user of the information); and the communication subsystem (the storage and retrieval mechanism permitting effective transmission between the receiver and the source). (Meadow, 1967) The present report is concerned with the automation of communication subsystems by means of IR programming systems, i.e., collections of programs designed to cause computers to act as communication subsystems for complete IR systems.

Since the present analysis places a major emphasis upon Generalized IR Programming Systems, some comment regarding the concept of system generality is necessary. Generality is an important characteristic of all problem oriented programming systems, whether designed for statistical analysis or information retrieval. Generality is typically defined in terms of the range of operations performed by a system. For our purposes, we shall define generality as a combination of adaptability and portability. Adaptability refers to the degree to which a system may be adapted to a variety of information processing problems. Portability refers to the ease with which the system may be implemented at various computer installations. Portability reflects both the machine independence of a system and the adequacy of its documentation.

Although the relative merits of generalized versus special purpose programming systems are frequently debated by computer scientists and programmers, current applications programming is clearly emphasizing generalized systems. The most probable reason for this is the average scientist's need

for the easier access to the computer provided by generalized systems. A familiar example of this trend is the development and rapid acceptance by behavioral scientists of generalized systems for statistical analysis--such as the Cooley and Lohnes (1962) system, the Dixon (1964) BMD System, and the BCTRY System (Tryon and Bailey, 1966).

The trend toward generalized systems is also clearly apparent in IR programming. The development of generalized IR programming systems appears to be a rather obvious extension of two major areas of prior and concurrent program development. One basis for generalized systems is provided by programs designed to sort and merge alphabetic and numeric information. The best known examples are the Key-Word-In-Context (KWIC) programs which produce printed indexes by permuting key words selected from author-title abstracts (Luhn, 1966). A second basis for generalized systems is provided by the specialized automated searching programs developed for particular applications--principally in the defense related industries. Currently, generalized IR programming systems formed by combining printed indexing and automated searching programs appear to be in a rapid stage of development. A recent federal survey of nonconventional scientific and technical information systems lists 118 computer systems (NSF, 1966b). Of these, 20 produce only printed indexes; 17 produce only automated searches; but 81 produce both printed indexes and automated reference searches.

The following table summarizes several typical generalized IR programming systems currently available for use by educators and behavioral scientists.

TABLE 2 Representative IR Programming Systems

SYSTEM	SOURCE	RESUME
GIS, Generalized Information System	IBM (1965) IBM Corp.	General purpose file maintenance and searching system with full information algebra. Record rather than document oriented. Assembly Language, IBM 360 Series.
INFOL Information Language	Olle (1967) CDC (1966) Control Data Corp.	General purpose file maintenance and searching system with limited information algebra. Record oriented. Assembly Language, CDC 3000 Series.
GPFS, General Purpose FORTRAN System	Bentley & Katz (1967) Sperry Rand	Limited purpose file maintenance and searching system with limited information algebra. Record oriented. FORTRAN. Limited to CDC 3000 Series due to dependence on SCOPE monitor.
General Inquirer	Stone (1962) Harvard University	Linguistic content analysis system. Given a dictionary thesaurus and text, system generates word frequency statistics for terms encountered in text. Limited retrieval capability.
Rand Catalog	Kay & Ziehe (1965) Rand Corp.	Very flexible general purpose file maintenance system designed for linguistic analysis. Permits definition of "data structures" linking information for retrieval. Assembly Language.
TRIAL, Technique for Retrieving Information from Abstracts of Literature	Janda (1966) Northwestern University	General purpose document retrieval system, without information algebra or linguistic analysis. Designed to retrieve abstracts in response to queries submitted by users. Assembly Language, IBM 7000 Series.
BIRS, Basic Indexing and Retrieval System	Vinsonhaler (1967) Michigan State University	General purpose document retrieval system without information algebra. Designed to perform automatic content analysis, storage, and retrieval. FORTRAN IV for CDC 3000 6000. Modifiable for IBM 7000 and 360 Series.

Several useful comments may be made regarding the contents of the table. First, there appear to be several major lines of development. The generalized system approach is best exemplified by GIS. This system is designed to be maximally adaptable. The user may make use of a complete "information algebra" permitting remarkable flexibility in communicating with his information file. The difficulties involved in generating such completely adaptable systems are indicated by the fact that GIS remains incomplete at the present writing. However, less general systems, such as INFOL and GPFS, are currently available. The research system approach is illustrated by the General Inquirer and the Rand Catalog. Both systems permit a specialized type of analysis aimed at a particular field of research--in this case linguistic analysis. The document retrieval system line of development is illustrated by the last two systems summarized in the table. Both TRIAL and BIRS are designed to store and retrieve documents in the behavioral sciences. Second, the design of current systems appears to emphasize adaptability over portability. Thus, the most adaptable systems tend to be the least portable. For example, GIS, INFOL, and the Rand Catalog are all written in the assembly language of particular computer systems. Third, a significant collection of general purpose IR programming systems is currently available for use in the behavioral sciences.

Scientists embarking upon research projects requiring the development of IR systems are well advised to examine the literature for relevant programming systems. In this regard, educational and behavioral science journals offer only limited indications of current IR programming. More complete summaries are available in general reviews by agencies of the

federal government (Berul, 1964; Fossum, 1966; NSF, 1966a; NSF, 1966b) and in recently published books in the field of information systems (Strauss, et al., 1966; Carter, et al., 1967; Schecter, 1967; Meadow, 1967).

III. BASIC INDEXING AND RETRIEVAL SYSTEM (BIRS)

A. PURPOSE

The Basic Indexing and Retrieval System is a set of fundamental programs designed to allow scholars and scientists to use their own locally based computer to construct and maintain a variety of IR Systems. Thus, BIRS may be viewed as a set of essential IR tools. The research worker may use these tools to construct the type of IR System which best meets his immediate needs.

In general, BIRS may be adapted to the development of any IR System having the following characteristics. First, each of the informational elements (e.g., documents, summaries of documents, references to documents, etc.) retrieved by the IR System must be punched on cards. Second, for purposes of indexing or retrieval, each of these informational elements must be describable by a collection of terms (e.g., a set of key-words indicating the topical content of the element, etc.). Within these rather broad limits, BIRS may be used to construct and maintain a variety of IR Systems ranging from computer generated books with author/subject indexes, to computer based searching and retrieval systems with automated retrieval of informational elements in response to users' questions.

The development of BIRS has particularly emphasized three major characteristics which have proved especially important for general purpose systems: portability, simplicity, and adaptability. Portability refers to the ease with which BIRS, or IR Systems based upon BIRS, may be transferred from one computer installation to another. In general, such systems

should be easily implemented at most university computer centers, since BIRS is written entirely in a FORTRAN IV level language, and uses the minimum storage configuration generally available in scientific computer centers: 32,000 words of core storage and six tape drives. BIRS was developed and tested on the Control Data Corporation 3600 at Michigan State University. Implementations of the system at other installations are now in progress.

Simplicity refers to the ease of using BIRS or IR Systems based upon BIRS. In order to be of maximal value in education and behavioral science, BIRS should be easily used by any researcher, regardless of his technical background in IR or computer science. Accordingly, the system is constructed to handle IR processing in the most straightforward manner with a minimum degree of control required of the user. In addition, a comprehensive Technical Manual (Vinsonhaler, 1967) has been developed for the system which provides a background reading in IR, and thorough descriptions of all programs.

Adaptability refers to the degree to which individual scholars may adapt the system to their own purposes. This characteristic is of particular importance in educational and behavioral science IR, because of the diversity of informational needs and the necessity for experimentation with various types of IR methods within these fields. Adaptability implies both ease of modification and flexibility of use. Ease of modification is designed into the sytem by using so-called "open-ended, modular" construction and by organizing the programming into independent clusters of dependent operations. Thus, each program in BIRS performs a separate basic operation and may be modified or replaced without major changes in the remaining components of the system. Flexibility is built into the system by (1) completely

automating only those clerical operations common to most IR systems; (2) allowing the user to control the operations which are nonclerical and dependent upon his particular IR needs; and (3) providing the user with a choice among programs and among methods within a single program.

B. THEORY

Most currently functioning IR programming systems are designed to implement rather specific types of IR Systems. Hence, such systems are developed ad hoc with minimal theoretical foundations. In contrast, BIRS is designed to implement many types of IR Systems. Consequently, an attempt has been made to establish an explicit theoretical basis for the methods used in developing the system.

The general design of BIRS is based upon an extension of Vickery's (1965) theoretical analysis of IR Systems. Essentially, this theoretical model is an abstraction of the functioning of most traditional libraries, and provides a basis for systems like BIRS, which permit users to develop automated or partially automated IR Systems, which are firmly based upon familiar traditional-library methods.

The fundamental theoretical model for BIRS is summarized in Table 3 and is discussed at length in the Technical Manual for the system. A complete discussion of analogous IR System models is available (Fossum, 1966).

The BIRS theoretical model assumes three types of fundamental informational elements: abstracts, descriptions, and questions. Three fundamental operations are associated with these elements: information storage,

indexing, and retrieval. These basic operations may be clarified by analogy with traditional library operations. Information storage corresponds to the maintenance of the collection of books and periodicals in the traditional library, i.e., the storage of documents for retrieval by access label or "call number." Information indexing corresponds to the maintenance of an index to the contents of the library, i.e., the maintenance of author, title, or subject catalogs. Finally, information retrieval operations are analogous to the process of searching author-title, or subject catalogs for the "call numbers" of relevant documents, and then retrieving the documents, themselves, by means of their "call numbers."

TABLE 3: Fundamental Operations Performed by IR Systems

OPERATION	DESCRIPTION
INFORMATION STORAGE	These operations are concerned with the preparation and storage of the basic informational elements to be processed by the system.
ABSTRACTING	Abstracts (a) are defined as basic informational elements, e.g., summaries of documents. This process consists of the preparation of abstracts for input to the IR System.
INFORMATION FILE MAINTENANCE	Information files are defined as collections of abstracts. The contents of the file are sets of the form $<L_a,a>$, where a denotes any abstract and L_a denotes a unique label or access number for the abstract. This operation consists of preparing or updating the file, so that abstracts may be located by means of the unique access labels.
INFORMATION INDEXING	These operations are concerned with the preparation and storage of descriptions of the topical contents of abstracts.
ANALYSIS	Descriptions (d_a) are defined as topical descriptions of abstracts, which are to be used to locate abstracts by subject content, e.g., sets of key-words. Analysis is concerned with the preparation of descriptions for each abstract in the Information File.
DESCRIPTION (INDEX) FILE MAINTENANCE	Description files are defined as indexes to information files, i.e., collections of descriptions and labels of the form $<d_a,L_a>$. These files provide a link between topic descriptions (d_a) and abstracts (a) by means of the unique access labels. The operations consist of preparing or updating the contents of the index.
INFORMATION RETRIEVAL	These operations are concerned with the retrieval of abstracts stored in the information file $<L_a,a>$, by means of descriptions stored in the index or description file $<d_a,L_a>$, in response to questions (q) submitted by users.
QUESTIONING	Questions are defined as requests for abstracts by topic. A question (q) is a description of a class of abstracts, e.g., a set of key terms defining the type of information of interest. Questioning refers to the process of stating such search requests.
SEARCHING	Searching refers to the process of using the description file to locate relevant abstracts. Thus, searching includes reading questions (q), comparing questions and descriptions (q,d_a), and storing the locations (q,L_a) of relevant abstracts.
RETRIEVING	Retrieving refers to the process of reading relevant abstracts from the information file, i.e., using the results of searching (q,L_a) to output relevant abstracts, (q,a).

Essentially, BIRS is an open-ended collection of independent computer programs organized according to this general theoretical model. Thus, the system includes a separate set of computer programs for each of the three fundamental operations of information storage, indexing, and retrieval. Each component program is designed to aid the user with a single fundamental operation. To construct a particular type of IR System, the user simply selects the proper component programs. The main advantage of this "modular" design is flexibility. Since the fundamental operations are mutually independent, the component programs may be used or modified independently. Hence, different components may be combined in various ways to produce a variety of IR Systems. Similarly, the entire system may be selectively modified by replacing existing components (or adding new ones) for one fundamental operation--without any modification of components designed for other basic operations.

The fundamental operations performed by BIRS are summarized in Table 4. As indicated in the table, the preparation of abstracts and questions must be performed by the user. All other fundamental operations may be performed by the user, performed by the user aided by BIRS, or performed entirely by BIRS, under the control of the user.

TABLE 4: Fundamental IR Operations Performed by BIRS

OPERATION	PERFORMED BY USER	AIDED BY BIRS	PERFORMED BY BIRS
Information Storage			
Abstracting	Yes	No	No
Information File Maintenance	Yes	Yes	Yes
Information Indexing			
Analysis	Yes	Yes	Yes
Description File Maintenance	Yes	Yes	Yes
Information Retrieval			
Questioning	Yes	No	No
Searching	Yes	Yes	Yes
Retrieving	Yes	Yes	Yes

C. BIRS/I

The initial version of the Basic Indexing and Retrieval System (BIRS/I) is now fully operative on the CDC 3600 at Michigan State University. Essentially, BIRS/I is designed to generate IR Systems using the so-called method of "coordinate index"--descriptions of abstracts are limited to unordered sets of key descriptive terms.

The computer programs included in BIRS are each designed to perform one of the three basic operations of IR Systems described in the previous section. Table 5 summarizes the functions performed by the seven programs included in the initial system. Briefly, BIRS/I includes a minimal collection of component programs covering the three major operations. Thus, one program (IFMP) is provided to aid the user with information storage; three programs (DAP, DFMP, PIP) are provided for information indexing; and two (DFSP, IFRP) are provided for information retrieval. Finally, one additional program (EXEC) is provided to aid with the manipulation of the system as a whole, e.g., to call components into operation from the system library tape.

In constructing any particular IR System with BIRS, the researcher may utilize various types of cards, tapes, and printed reports. The cards, tapes, and reports used or generated with the system may be classified according to the fundamental IR System operation which they serve.

TABLE 5: Component Programs of the Basic Indexing and Retrieval System (BIRS/I)

OPERATION	PROGRAM	FUNCTIONS
Executive	EXEC	The Executive Program is designed to store and retrieve component programs comprising BIRS.
Information Storage	IFMP	The Information File Maintenance Program is designed to read informational elements from cards, assign a unique access number to each element, and store the element on the Information File Tape (IFT), so that the element may be retrieved, given its access number.
Information Indexing (Descriptive Analysis)	DAP	The Descriptive Analysis Program is designed to aid the user with the task of indexing or classifying informational elements. In general, DAP reads informational elements either from the IFT or from cards and searches them for key words. Thus, DAP may be used either to perform a word frequency analysis or to automatically index informational elements.
Information Indexing (Description File Maintenance)	DFMP	The Description File Maintenance Program is designed to read descriptions (i.e., sets of indexing terms) and access numbers of informational elements from cards and store them on the Description File Tape (DFT) to provide an index to the contents of the Information File Tape (IFT). The user may manually generate the card input for the DFMP, or use either the IFMP or the DAP, to aid him with this task.
Information Indexing (Printed Indexing)	PIP	The Printed Indexing Program is designed to prepare a traditional author or subject index using informational elements read from cards or from the IFT, or using descriptions read from cards or from the DFT. The IFMP may be used in conjunction with this program to generate a listing of informational elements organized by access number.
Information Retrieval (Automated Searching)	DFSP	The Description File Searching Program is designed to read requests for particular types of informational elements (stated as sets of key terms) from cards, to search the DFT for relevant informational elements, and to store the access numbers of the most relevant elements on the Question File Tape (QFT).
Information Retrieval (Automated Retrieval)	IFRP	The Information File Retrieval Program is designed to read requests and access numbers of the relevant informational elements from the QFT, retrieve the corresponding informational elements from the IFT, and print both the requests and the relevant informational elements for the user.

Table 6 summarizes the main card input for IR Systems developed with BIRS/I. As shown, abstracts, descriptions of abstracts, and questions are all input on cards. The abstract and question cards must be prepared by the user. The descriptor cards may be prepared by the user or by BIRS programs.

It should be noted that not all of these types of cards are needed for all applications.

Table 7 summarizes the tapes which may be used with IR Systems developed by BIRS/I programs. As with component programs and cards, each tape is concerned with one of the major IR operations: the IFT contains abstracts, the DFT contains descriptions of abstracts, and the QFT contains questions and the access numbers of abstracts relevant to the questions.

As before, not all IR Systems will require the use of all tapes. In general, only completely automated systems require all tapes. Partially automated systems replace tapes with printed reports.

Table 8 summarizes the printed reports used with IR Systems developed from BIRS/I programs. As indicated, each type of report is mainly concerned with one of the three major operations.

As before, different types of IR Systems will require the use of different types of printed reports. Thus, systems emphasizing manual searching and retrieval will require the use of a printed information file and a printed index, while automated searching and retrieval systems will require printed reports containing abstracts relevant to specific questions.

To summarize the preceding discussion, BIRS/I is essentially a general purpose indexing and retrieval system. It is designed to permit the construction of IR Systems which locate summaries of more complete documents stored in auxiliary files. Thus, IR Systems developed with BIRS are analogous to the reference services offered by traditional libraries. Thus, BIRS/I programs may be used to produce a printed equivalent of the author-title, subject index or an automated approximation to the reference-librarian.

TABLE 6: Cards Used with Component Programs of BIRS/I

OPERATION	CARD TYPE	FUNCTION
Executive and Control	BIRS Control Cards	Control cards are used to select component programs and govern their operations. Each card is punched with a *$ in the first two columns, followed by a control phrase. For example, cards calling the IFMP would contain *$IFMP.
Information Storage and Indexing	Abstract Cards	Abstract cards are used to input informational elements to be processed by the system. An abstract may consist of any set of information punched on from one to fifty cards. Each abstract is identified by a unique "access number" or "label," normally assigned by BIRS programs under the user's control. The following is an example of a very simple abstract punched on cards. *$ABSTRACT "LABEL" BENNETT, G.K., BENNETT, M.G., WALLACE, W.L., AND WESMAN, A.G. COLLEGE QUALIFICATION TEST. PSYCHOLOGICAL CORP. GENERAL INTELLIGENCE FOR COLLEGE ENTRANTS AND ADULTS. SIXTH MENTAL MEASUREMENTS YEARBOOK, O.K. BUROS, GRYPHON PRESS 1965, ENTRY NO. 450. TEST LIBRARY NO. C1033.
Information Storage and Indexing	Descriptor Cards	Descriptor cards are used to input topical descriptions of abstracts for further analysis. Descriptions are sets of key terms describing the topical content of abstracts. Each abstract is separately described by a set of up to 30 terms, identified by means of the "label" or access number of the abstract being described. Terms are separated by commas, and the latter portions of long terms are ignored. There are no other restrictions on the content of terms. Descriptor cards may be generated by the user or by BIRS/I programs. The following is an example of a description of the abstract given above. *$DESCRIPTOR "LABEL" COLLEGE, APTITUDE, VERBAL, QUANTITATIVE, ADULTS
Information Retrieval	Question Cards	Question cards are used to query automated IR Systems developed with BIRS/I, i.e., to request that component programs search for and retrieve abstracts relevant to given topics. Question cards are used to indicate the topics of interest. Each search request is indicated by means of a set of up to 30 terms, separated by commas. The following is an example of a request for abstracts dealing with general intelligence for adults. *$QUESTION GENERAL, INTELLIGENCE, ADULTS

TABLE 7: Tapes Used with Component Programs of BIRS/I

OPERATION	TAPE	CONTENTS AND FUNCTION
Executive	BIRS Library	The BIRS Library Tape contains component programs of the system. The programs are read from cards and stored on the tape by the Executive Program. Programs may be called from the tape by means of executive control cards.
Information Storage and Retrieval	IFT	The Information File Tape contains abstracts read from cards and stored so that each can be located given its unique access number or "label." The IFT is generated by the IFMP and is used by the DAP, the PIP, and the IFRP.
Information Indexing and Searching	DFT	The Description File Tape contains key-term descriptions and access numbers or "labels" of abstracts. Descriptions and "labels" are associated so that abstracts may be located by topical content. The DFT is generated by the DFMP from descriptor cards prepared by the user, or by the DAP. The tape is used by the PIP, and the DFSP.
	QFT	The Question File Tape contains questions read from question cards, and access numbers of relevant abstracts. The QFT is generated by the DFSP and is used by the IFRP.

TABLE 8: Printed Reports Used with Component Programs of BIRS/I

OPERATION	REPORT TYPE	CONTENTS AND FUNCTION
INFORMATION STORAGE	Printed Information File	This report consists of a printed equivalent of the Information File Tape, i.e., a listing of abstracts printed in numerical order of access number. The printed Information File is generated by the IFMP from abstract cards or from the IFT.
INFORMATION INDEXING	Content Analysis Report	This report consists of a content analysis of word usage in abstracts, read either from cards or from the IFT. The report is generated by the DAP, which also may be used to automatically index each abstract by selecting key terms and punching them on descriptor cards.
	Printed Index	This report consists of a printed equivalent of the Description File Tape, i.e., an index for locating abstracts by topic. Since this index is designed for manual searching, a term-entry index is prepared. That is, all key-terms are printed in alphabetical order and each term is followed by the "labels" or access numbers of abstracts relevant to the term. The printed index is produced by PIP from abstract cards, the IFT, descriptor cards, or the DFT.
INFORMATION RETRIEVAL	Abstracts relevant to questions	This report consists of the results of an automated search performed by the DFSP. Each question originally stated on cards is printed and following it the relevant abstracts--in order of relevance. The report is generated by IFRP from the QFT and IFT.

D. BIRS/II

The second version of the Basic Indexing and Retrieval System (BIRS/II) is to be released for national distribution (with a complete Technical Manual) in December, 1967. BIRS/II is identical to BIRS/I except for the following extentions: (1) Coordinate and weighted indexing--terms or terms and numerical weights may be used to describe documents; (2) Word root analyses and synonym dictionaries--terms may be reduced to generic roots, and synonyms may be proved during indexing and/or searching; and (3) Generalized logical searching--user may call upon simplified relevance searching or sentential logic searches for question answering. Table 9 summarizes the functions of the ten computer programs comprising BIRS/II.

The cards, tapes, and printed reports used with the component programs of BIRS/II are identical to those used with the initial system except for the addition of facilities permiting weighted indexing, word root analysis, synonym dictionaries, and logical searching. Table 10 summarizes the major types of additional card input available in BIRS/II.

TABLE 9: Component Programs of the Basic Indexing and Retrieval System

FUNDAMENTAL IR OPERATION	COMPONENT PROGRAM	FUNCTIONS OF COMPONENT PROGRAM
Executive	EXEC	The Executive Program is designed to store and retrieve component programs comprising BIRS.
Information Storage	IFMP	The Information File Maintenance Program is designed to read informational elements* from cards, assign a unique access number to each element, and store the element on the Information File Tape (IFT), so that the element may be retrieved, given its access number.
Information Indexing (Descriptive Analysis)	DAP	The Descriptive Analysis Program is designed to aid the user with the task of indexing or classifying informational elements. In general, DAP reads informational elements either from the IFT or from cards and searches them for key words. Thus, DAP may be used either to perform a word frequency analysis or to automatically index informational elements
	FAP	The File Analysis Program is a highly generalized content analysis program which operates on natural language text. FAP performs the same operations as DAP in the indexing process.
Information Indexing (Description File Maintenance)	DFMP	The Descriptive File Maintenance Program is designed to read descriptions (i.e., sets of indexing terms) and access numbers of informational elements from cards and store them on the Description File Tape (DFT) to provide an index to the contents of the Information File Tape (IFT). The user may manually generate the card input for the DFMP, or use either the IFMP or the DAP, to aid him with this task.
Information Indexing (Printed Indexing)	PIP	The Printed Indexing Program is designed to prepare a traditional author or subject index using informational elements read from cards or from the IFT, or using descriptions read from cards or from the DFT. The IFMP may be used in conjunction with this program to generate a listing of informational elements organized by access number.
	PLP	The Printed Listing Program is designed to provide printed books, i.e., listings of abstracts, ordered by the contents of the abstracts, i.e., principal author's last name, etc. The printed books produced by PLP are similar to those produced by IFMP, except that the books of the latter are ordered by IFT access number.

*Documents or abstracts of documents not exceeding 50 cards in length

TABLE 9: (continued)

FUNDAMENTAL IR OPERATION	COMPONENT PROGRAM	FUNCTIONS OF COMPONENT PROGRAM
Information Retrieval (Automated Searching)	DFSP	The Description File Searching Program is designed to read requests for particular types of informational elements (stated as sets of key terms) from cards, to search the DFT for relevant informational elements, and to store the access numbers of the most relevant elements on the Question File Tape (QFT).
	DFSPL	The Description File Logical Searching Program is similar to the DFSP except that it is designed to perform logical searches.
Information Retrieval (Automated Retrieval)	IFRP	The Information File Retrieval Program is designed to read requests and access numbers of the relevant informational elements from the QFT, retrieve the corresponding informational elements from the IFT; and print both the requests and the relevant informational elements for the user.

TABLE 10: Additional Cards Used With Component Programs of BIRS/II.

CARD TYPE	FUNCTION
Synonym Cards	Inputs the thesaurus or synonym dictionary to be used in indexing or searching. The following example equates "ability" and "aptitude" with "intelligence." *$ SYNONYMS INTELLIGENCE = ABILITY, APTITUDE
Descriptor Cards	A new form of the descriptor card permits the use of numerical weights to indicate the degree of importance for indexing terms. The following example indicates that 75% of the items in the described test measure verbal ability, while only 25% of the items measure quantitative ability. *$ DESCRIPTION VERBAL = .75, QUANTITATIVE = .25, INTELLIGENCE = 1.00
Question Cards	The question cards used with BIRS/II may optimally include relevance weights or logical operators to further control the automated search. The first example indicates a relevance search with an emphasis upon tests measuring verbal rather than quantitative ability. The second example illustrates a typical logical search excluding quantitative ability tests. *$QUESTION INTELLIGENCE = 1.00, QUANTITATIVE = .25, VERBAL = .75 *$QUESTION (INTELLIGENCE OR APTITUDE) AND (NOT QUANTITATIVE).

IV. APPLICATIONS

The Basic Indexing and Retrieval System (BIRS/I and BIRS/II) has been very well received by educators and behavioral scientists.

The various reports regarding the system have generated a substantial correspondence which includes requests for further information, orders for complete documentation, and numerous pledges of cooperation in implementing the system at diverse computer installations.

We have no way of knowing all of the applications currently being made with BIRS. The major applications familiar to the present author are summarized in Table 11. These applications are at best representative; they do not even include all of the BIRS applications at Michigan State University.

TABLE 11. Representative Applications of BIRS/I and BIRS/II

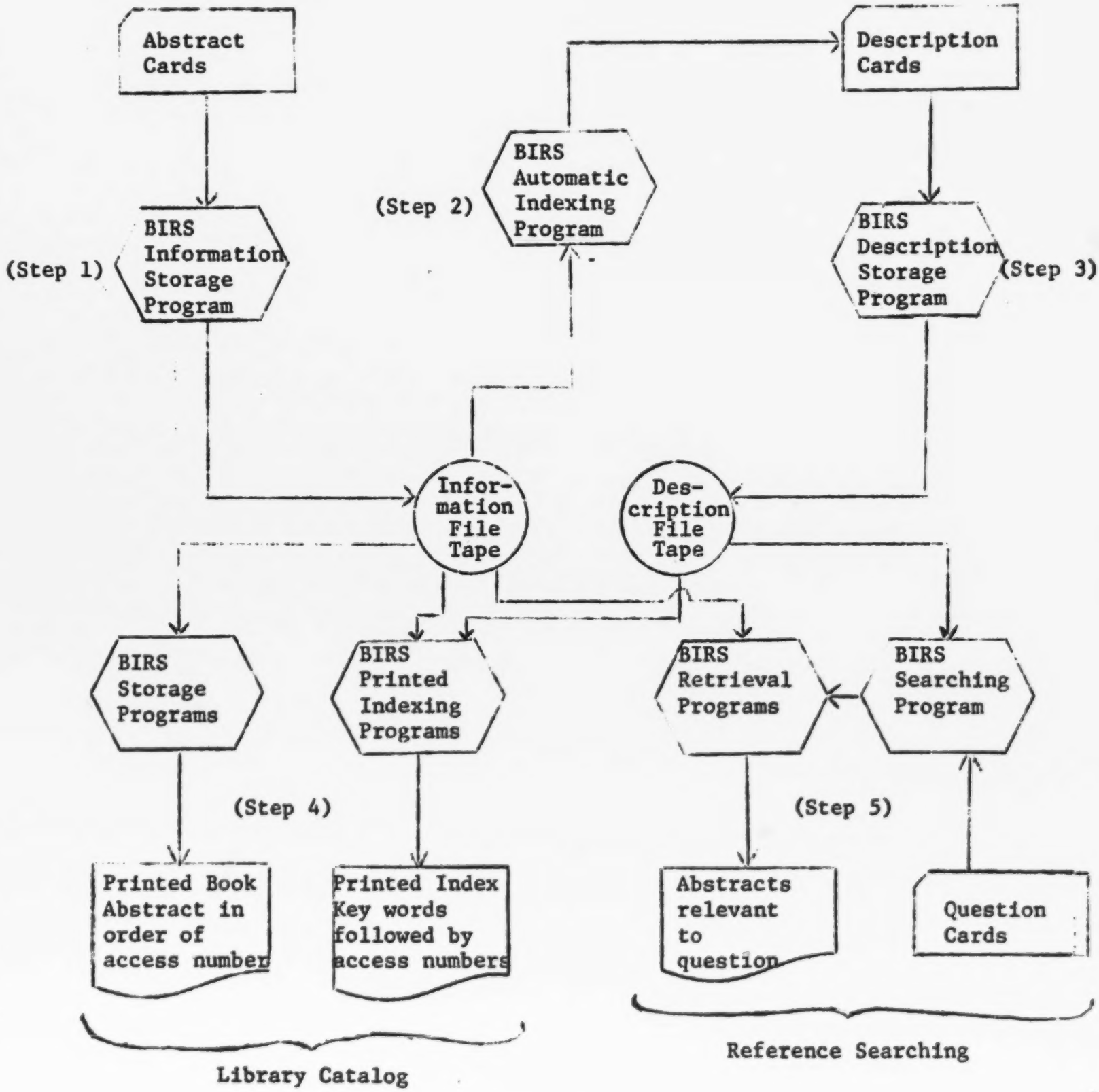
PROJECT	BIRS APPLICATION
Instructional Materials Center for Handicapped Children and Youth; MSU/USOE. Alonso and Ward, Principal Investigators	Automate Regional Library for materials center. Computer generated library catalogs, reference searching etc. on CDC 3600.
Instructional Materials Center. University of Texas/USOE:HCY. W. Wolfe, Principal Investigator.	Computer generated indexes for materials center collection, on CDC 6600.
Single Concept Film Clip Project. MSU/USOE. Prof. Woodrow Miller, Principal Investigator.	Computer generated indexed book for 850 single concept films, on CDC 3600.
ERIC Council on Exceptional Children Clearinghouse. Dr. June Jordan, Principal Investigator.	Preparation of accession lists for 15 clearinghouses and searches for duplication of abstracting among clearinghouses, on CDC 3600.
TEST Information Retrieval System Project. Vinsonhaler, Principal Investigator.	Automated indexing and retrieval for psychological tests relevant to aptitude measurement.
Michigan Inter-University Committee on Information Systems, (MICIS); jointly supported by University of Michigan, Michigan State University, Wayne State University, and the State of Michigan.	Automated preparation of a continually updated bibliography for computer applications in university education, including IR, CAI, and general applications of computers in university environments.

To clarify the nature of current BIRS applications, let us consider a particular example chosen from Table 11. The Instructional Materials Center for Handicapped Children and Youth is a regional library for documents relevant to the education of visually handicapped children. Presently, the library includes about 6,000 major documents, e.g., psychological tests, text books, periodicals, instructional devices, etc. All of the indexing and retrieval functions of the library are performed by BIRS, implemented on the university CDC 3600 System.

The procedures are as follows. As new documents are acquired, they are assigned an arbitrary accession number and stored in the library. Next, an abstract is prepared for each document which includes bibliographic information, a summary, and the document accession number. The abstract is punched on cards for input to BIRS.

Like most computer based libraries, the Materials Center provides users with both printed indexing services and automated reference searching. Thus, printed indexed books are provided in place of the traditional card catalogs while automated searching is provided in place of traditional reference service. The Materials Center staff uses BIRS mainly to prepare printed catalogs and to perform reference searches requested by users of the library. The procedures are summarized in the following figure.

FIGURE 1. Information Storage, Indexing, and Retrieval in a Typical Research Library



Normally, the processing begins with the storage of new abstracts on an information file tape (step 1 in the figure). Next, each abstract is indexed or classified and description cards are prepared which contain the indexing terms. Usually, the indexing is automated (as shown in step 2 of the figure). Finally, the description cards are input, and an index is constructed on a description file tape (step 3). Printed catalogs (step 4) are generated by programs which read abstracts and descriptions from the tapes. These programs then prepare books of abstracts and various types of key word indexes, e.g., by author, title, subject, etc. Reference searches (step 5) are obtained by inputting question cards describing topics of interest to one of the BIRS searching programs which use the description file tape to locate relevant abstracts. The BIRS retrieval program reads the relevant abstracts from the information file tape and prints them for the user.

Obviously, IR programming systems may be used for many purposes besides automating research libraries. For example, BIRS is currently being applied to produce reading lists for university courses, to prepare bibliographies for computer applications in education, to investigate methods of automated indexing and searching, and to perform content analyses of protocols for research in psychology. However, it must be admitted that the predominant usage of BIRS (and other similar systems) is in the automation of research libraries.

V. CONCLUSIONS

Several obvious conclusions may be offered with respect to the BIRS project. First, there appears to be a very real need for generalized IR programming systems in behavioral science and education. The success of BIRS probably stems from its provision for gradual steps beyond traditional library technology by the average university scientist or scholar.

Second, our experience indicates that computerized IR Systems are substantially less expensive than manual systems for specialized libraries. For the 6000 document Instructional Materials Center collection, Computer Generated Catalogs (with author/subject/indexes) cost about $120 per original copy, while automated reference searches cost about $ .60 per search. One reason for this low cost (in contrast to that of other computerized IR Systems) stems from the direct participation of non-technical personnel in applying BIRS.

The cost equations developed on the CDC 3600 (and probably applicable to other implementations) are as follows:

$$\text{\$ Cost for KWIC Index} = .002\ (\text{Number of Abstracts}) \times (\text{Average Number of Cards Per Abstract})$$

$$\text{\$ Cost for Each Question} = .0001\ (\text{Number of Abstracts})$$

The second factor in the first equation is the overall number of cards per abstract which are actually used in preparing the index.

Third, perhaps the most significant finding of the present project is that technology rather easily outdistances understanding. Thus, we have found that many of the extentions of BIRS/I, included in BIRS/II, are not

applied by users, because they lack a sound theoretical basis for such applications. Most of our users view elementary content analyses (e.g., word root analyses or synonym definition) with exaggerated suspicion.

Further, the practical application of certain facilities--such as probabilistic or weighted indexing (Maron & Kuhns, 1960)--are as much a puzzle to us as to our users. In short, we have come to the conclusion that further major developments in the technique of IR Programming Systems must necessarily be accompanied by careful study of the psychological or behavioral aspects of Information Retrieval and Communication.

VI. REFERENCES

Bentley, J. E. & Katz, S. M. A General Purpose Fortran System for File Maintenance, Retrieval, and Formatting. Fourth Annual National Colloquium on Information Retrieval, May, 1967.

Berul, L. Information Storage and Retrieval: A State of the Art Report. Auerbach Corp., September, 1964.

Bourne, C. P. Methods of Information Handling. Wiley, 1963.

Buros, O. K. The Sixth Mental Measurements Yearbook. Gryphon Press, 1965.

Buros, O. K. Tests in Print. Gryphon Press, 1961.

Carter, L. F. et al. National Document-Handling Systems for Science and Technology. Wiley, 1967.

CDC. Computer Systems INFOL, 3400-3600, General Information Manual. Control Data Corporation, Palo Alto, California, 1966.

CDC. Computer Systems INFOL, 3600 - 3800 Reference Manual. Control Data Corporation, Palo Alto, California, 1966.

Cooley, W. W. & Lohnes, P. R. Multivariate Procedures in the Behavioral Sciences. Wiley, 1962.

Dixon, W. J. (ed.). Biomedical Computer Programs. Health Sciences Computing Facility, School of Medicine, University of California at Los Angeles, 1964.

Fossum, E. G. et al. Optimization and Standardization of Information Retrieval Language and Systems. Clearinghouse for Federal and Technical Information, AD630 797, 1966.

Janda, K. & Tetzloff, W. H. "Trial: A Computer Technique for Retrieving Information from Abstracts of Literature." Behavioral Science, 1966, 11, 480-86.

IBM. Generalized Information System, Application Description. International Business Machine Corp., White Plains, N. Y., 1965.

Kay, M. & Ziehe, T. The Catalog: A Flexible Data Structure for Magnetic Tape. Rand Memorandum RM-4645-PR. Rand Corp., Santa Monica, California, 1965.

Luhn, H. P. "Keyword-in-Context Index for Technical Literature (KWIC Index). in Hayes, D. G. (ed.) Readings in Automatic Language Processing. Elsevier, 1966.

Maron, M. E. and Kuhns, J. L. "On Relevance, Probabilistic Indexing, and Information Retrieval." Journal of the ACM 1960, 7, 216-244.

Meadow, C. T. The Analysis of Information Systems. Wiley, 1967.

NSF. Current Research and Development in Scientific Documentation, No. 14. National Science Foundation NSF-66-17, 1966a.

NSF. Non Conventional Scientific or Technical Information Systems in Current Use, No. 4. National Science Foundation NSF-66-24, December, 1966b.

Olle, T. W. "INFOL, A Generalized Language for Information Storage and Retrieval Applications," in Schecter, G. (ed.) Information Retrieval. Thompson, 1967.

Schecter, G. (ed.). Information Retrieval: A Critical View. Thompson Book Company, 1967.

Stone, P. "The General Inquirer." Behavioral Science, 1962, 7, 484-98.

Strauss, L. J., Strieby, I. M., & Brown, A. L. Scientific and Technical Libraries. Wiley, 1964.

Tryon, R. C. & Bailey, D. E. "The BC TRY Computer System of Cluster and Factor Analysis." Multivariate Behavioral Research, 1966, 1, 95-111.

Vickery, B. C. On Retrieval System Theory. Second Edition, N. Y., Butterworth, 1965.

Vinsonhaler, J. F. "BIRS: A System of General Purpose Computer Programs for Information Retrieval in the Behavioral Sciences." American Behavioral Scientist, 1967, 10, 12-23.

Vinsonhaler, J. F. (ed.). Technical Manual for the Basic Indexing and Retrieval System. College of Education, Michigan State University, East Lansing, Michigan, 1967.